

# The Research on Cloud Storage

Cai-Qun Wang, Liu-Tao Zhao, Shou-Xin Cao

Cloud Computing Department, Beijing Computing Center, BCC, Beijing, China

E-mail: wangcq@bcc.ac.cn, zhaolt@bcc.ac.cn, caosx@bcc.ac.cn

**Abstract**—This paper introduces the concept of cloud storage and cloud storage architecture. Then it demonstrates the cloud storage type including Block Storage, File Storage and Object Storage. Object Storage has the advantages of both the Block Storage and File Storage. Object Storage can deal with massive data and structured, unstructured and half-structured data. Therefore, we propose a cloud storage scheme based on Object Storage in this paper. This cloud storage solution turns out to be a system with high availability, high fault tolerance and extendibility. Finally, the future development trend of cloud storage was explored. (Abstract)

**Keywords**—component; Object Storage; Cloud Storage; Cloud Computing (key words)

## I. INTRODUCTION

With the development of the Internet and the data increases explosively, data storage has already been a problem we need to face. More and more users choose cloud storage to deal with data storage problems. Cloud storage is a new mode for data store. It provides a more low-cost, more reliable and safer storage service. It uses less resource and encapsulates the technology detail about how to implement the service. The user can get adequate computing and storage resource as long as accessing to the "cloud".

Cloud storage adopts cluster application technology, grid technologies and DFS (distributed file system) technology. It gathers a large variety of different types of storage devices in internet to work cooperatively to provide data storage service and data access service.

Cloud storage is a service mode. Compared to traditional storage solutions, cloud storage is scalable, cost-effective, and easy access. It has brought a lot of convenience for individuals and enterprises.

## II. RELATED TECHNOLOGIES

### A. Cloud Storage

Cloud storage is a virtual storage mode based on the cloud computing and is mainly used to deal with massive data storage and real-time call problems. Cloud storage is a service concept rather than storage devices. It refers to virtual storage and provides multiple users with access to a shared storage pool.

In cloud storage system, large numbers of cheap storage devices make up a storage center in which devices are connected to each other by their respective distributed file system. Cloud storage provides high reliability and high scalability service. Cloud storage system has five layers in architecture.

### 1) Infrastructure

Infrastructure is the most basic part of cloud storage. Storage devices refer to channel storage devices such as FC (Fiber Channel); IP storage device such as NAS (Network Attached Storage), ISCSI (Internet SCSI); storage device such as SCSI (Small Computer System Interface), SAS (Serial Attached SCSI), and DAS (Direct-Attached Storage). Storage devices distribute wildly and are considerable. They connect with each other by wide area network, Internet or FC network.

### 2) Storage management

Storage management includes the logical device virtualization management, status monitoring, fault multi-link redundancy management and hardware maintenance.

### 3) Metadata management

Metadata management is the core part of the cloud storage and is also the most difficult to achieve. Based on cluster, distributed file system and grid computing technology, the storage devices work collaboratively to provide a service that with higher data access performance.

CDN (Content Delivery Network) content distribution system and data encryption technology ensure that data stored in the cloud cannot be accessed by unauthorized users. In the meantime, cloud storage provides security and stability through a variety of data backup and disaster recovery technologies. It can ensure that data is not lost in cloud storage.

### 4) Storage overlay

Storage overlay is the most flexible part. Different cloud storage operating units can develop different application service interface to provide different applications according to actual business. It exposes simplified and standard data structures to service interfaces. For example, video surveillance applications platform, IPTV (Internet Protocol Television) and video-on-demand application platform, network hard drive reference platform, remote data backup application platforms.

### 5) Service interface

Service interface provides users with a common application interface standard. Authorized user can access cloud storage via a common interface services. Different operating units of cloud storage provide different access types and access methods.

### B. Storage Type

The most popular type of cloud storage is Block Storage, File Storage and Object Storage (Object-based Storage). Compared to Block Storage and File Storage, Object Storage

(Object-based Storage) is newer as network storage architecture in some content.

#### 1) Block Storage

Block Storage includes to DAS (Direct-Attached Storage, DAS) and SAN (Storage Area Network, SAN).

DAS is directly connected to host server, each host server has a separate storage device. Host server storage device cannot communicate with each other. It is relatively complex to access data across host. If host servers belong to different operating systems, it is more complex to access each other's data and some cannot even access the data. Thus, DAS is usually used in a single network environment and the amount of data exchanged is small.

SAN connects to network with professional host server through high-speed network. It locates at the end of the host group in which servers use high-speed I/O to link each other. Generally, SAN applications have high access speed, high data reliability, security and high performance for data sharing. It is preferred by high-performance computing because of high bandwidth and low latency. But SAN system has higher price and poor scalability. It cannot meet the demand with the number of CPU increases.

#### 2) File Storage

NAS (Network Attached Storage, NAS) is a set of network storage devices which are usually directly connected to the Internet and provide high cost performance and access service like a system. Usually, the NAS products provide file level storage service. It uses NFS (Network File System) or CIFS(Common Internet File System) commands to access data, transfer data as file and achieve network storage based on TCP/IP. So, File Storage has high scalability, cheap price and easy-management. But it is not suitable for high performance cluster because NAS's protocol is hard to implement and NAS has low bandwidth and big delay in data transfer.

#### 3) Object Storage

Generally speaking, Object Storage combines the high-speed direct access of SAN and distributed shared of NAS.

Data accesses (data read or write) channel and control (metadata) channel is separate in Object Storage system. It is build based on Object Storage devices (Object-based Storage Device, the OSD). Object Storage devices show certain intelligence in managing the distribution of data automatically. Object Storage combines the advantages of File Storage and Block Storage so that it can satisfy the users' need to store large amount of data and unstructured data.

### III. CLOUD STORAGE SYSTEM BASED ON OBJECT STORAGE

Object Storage combines the advantage of Block Storage and File Storage. It can store data no matter how huge it is and is compatible with a variety of data structures. Therefore, we adopt the Object Storage to build cloud storage system.

#### A. Storage model

The system can provide users with public storage service. Be different from traditional Block Storage, this system have a non-relational database in the back-end. The innovation

architecture design guarantees the efficiency when reading and writing unstructured data. This storage system can provide users with distributed storage service which is based on the object. Data in the storage system has multiple backup which is implement automatically. It improves the data availability. The data in the system is still available even if one storage node fails. This system is very suitable for storing unstructured data and satisfies users' higher requirements for data availability.

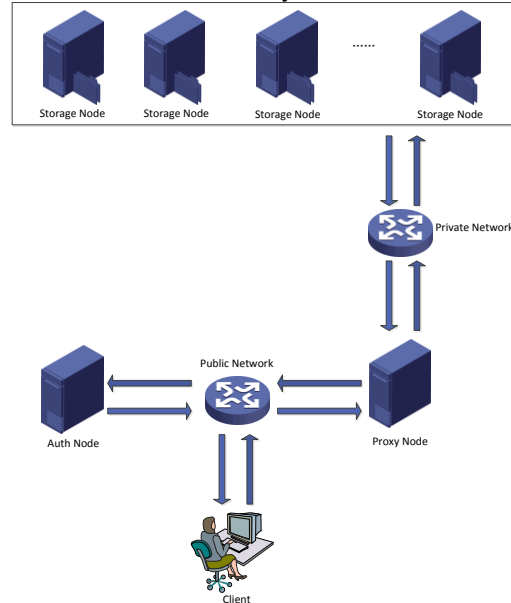


Figure 1. The Topology of Cloud Storage System

#### 1) Proxy Node

Proxy Node is the controller of Cloud storage system. It manages the system and indexes metadata and the actual data. It manages the high capacity storage and provides data redundancy function. It manages the back-end storage devices to provide high-performance concurrent access. Proxy Node server is the commander of the whole management platform in which administrators can monitor all operation, manage users and provide other useful strategies.

#### 2) Storage Node

Cloud storage system embedded cloud storage system access protocol and storage node license. It uses high-performance storage devices such as high-density disk array devices as infrastructure. Storage devices access cloud storage system via internet. Cloud storage system put the physical device into the virtual storage pool and allocates storage to users. The data store in the system has multiple copies and distribute in multiple physical devices. It increases the storage capacity by increasing the number of data node devices when storage capacity is insufficient. It is flexible to increase storage capacity according to the actual need. And it is free to increase capacity and improve performance when the system is running.

### B. Storage System based on Object Storage

Cloud storage system is based on distributed database system. It can provide storage services or be used to set up a cloud and private cloud. In the meantime, the system has API interface to be compatible with Amazon S3.

The picture below shows how to store a big object data in the cloud storage system. The object can be uploaded to the system via storage API. Cloud storage system divides the object into multiple fluxion-structure blocks. Then block is stored in the distributed database system in the back-end. Metadata is offered to users for retrieving data later.

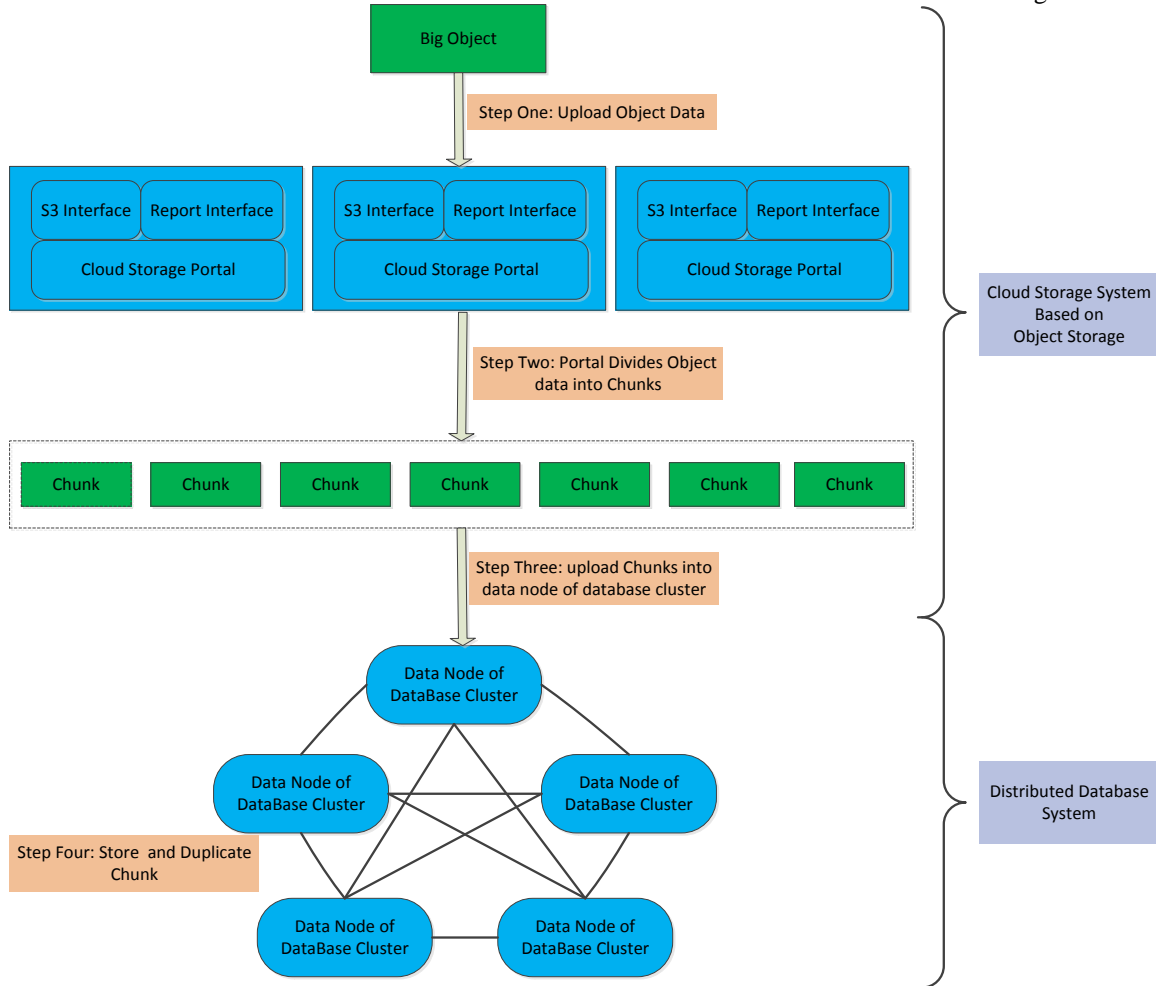


Figure 2. The Procedure to store a big Object Data

Data is stored in object storage system based on Key-Value model. The flat storage structure can greatly enhance efficiency in data query and storage. The system can provide multiple users with independent data storage service and safety at the same time. It is compatible with Amazon S3 and provides streaming division function for large data objects. You can get the storage service easily on the Web. The system has load balance part to prevent storage hot-spot problem.

### C. Backup for Disaster Recovery

The system also has master-slave disaster preparedness as Figure 3. It ensures the high availability by implementation anti-disaster backup through master-slave replication.

The system can satisfy disaster recovery and regulatory requirements through monitoring the data and master-slave

replication. The data in the system realizes real-time synchronization, full synchronization and synchronization in specific location.

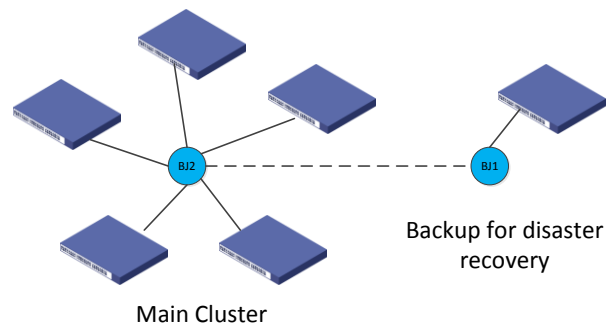


Figure 3. The Network Topology of Management System for Multiple Data Center

The cloud storage system greatly simplifies the developers' work on the data storage. Reasonable storage solution can guarantee the data safety and reliability. System developers have to specify servers and attachments manually when they store data in traditional mode. It is easy to cause server load imbalance because it is hard to effectively evaluate the data growth. And cloud storage system can balance the data. It automatically store data to different physical nodes. And it ensures that the data is not lost when a physical server fails. The cloud storage system provides users with high reliability of cloud storage service.

#### IV. THE FUTURE OF CLOUD STORAGE

Cloud storage has become a trend of the storage. With the development of cloud storage technology, vendors take the initiative to combine search technology, application technology and cloud storage with cloud storage to provide a range of data services. However, cloud storage will have a big improvement in security, portability and availability according to development trend.

##### A. Security

Security has always been one of the priority problems for enterprises to implement cloud computing from the very beginning. It's the same in cloud storage. Security remains a primary consideration for customers who want to use cloud storage. Security is usually the priority commercial considerations and technical considerations. Many cloud storage users demand higher level security than their own security architecture can provide. Facing such unrealistic high security requirements, many large reliable cloud storage vendors still try to meet their demand. They build much more secure data centers. It turns out cloud storage is with fewer security vulnerabilities and higher level security than the cloud storage users' data center.

##### B. Portability

Portability refers to data portability and the portability of cloud storage infrastructure.

Users will consider data portability when they want Managed Storage Services. Some large service providers promised to provide a solution with data portability like data is stored locally. Some cloud storage combines such a powerful portable that the whole data set can be transferred to any device you want even a specialized storage device.

In the view of IT infrastructure development, portability refers to infrastructure can move freely. Complex cloud storage system integration mode has become obstacle for business innovation. Simplify and optimize infrastructure can provide better protection for cloud storage. So cloud smart appliance appears. Smart appliance is an integrated system product which combines software and hardware system. It generally includes data processing part, data transmission part and data storage part. Smart appliance can deploy rapid, simplify the IT infrastructure, save resources and improve system's availability and scalability by integration, testing, and optimization in advance.

##### C. Performance

There is delay when accessing data in managed storage and remote storage over the past few years. The local cache is a big breakthrough for cloud storage. Data is kept local to alleviate the Internet latency issues. The delay problem can also be alleviated through the local cache even when we face the most serious network outages. Cloud storage can provide the same availability, performance, safety and visibility with NAS terminal device by a local NAS gateway. The manufacturers strive to achieve capacity optimization and WAN (wide area network) optimization to minimize data transmission delays with the development of cloud storage technology.

##### D. Reliability

One of the most popular methods to enhance the reliability of the data in storage system is the data redundancy backup. It can improve the reliability of the system. However, users will pay more for physical space, bandwidth with the increase of copy capacity. The system performance will also be impact performance negatively. How to guarantee the reliability more effectively will be one of the hottest areas of research in the future with the continuous development of cloud storage technology.

#### V. CONCLUSION

Compared to File Storage and Block Storage, Object Storage technology appears relatively late. It will become the mainstream storage technology in the future with the data cluster grows. The cloud storage system based on Object Storage has many advantages.

**High availability:** The system can guarantee the availability of data even when the individual servers or network fails.

**Simple Operation:** we only add to the cluster machines when we need more capacity.

**Extendibility:** The system can achieve load balance automatically when capacity scales up.

**Fault Tolerance:** The system will not result in data loss due to network partition or hardware failure.

With continuous development of Cloud Storage technology and merger of massive commercial cloud computing platform, Cloud Storage technology will get more attention in the near future. It will become the most widely used shared-platform and play a bigger role in wild fields.

#### REFERENCES

- [1] Chen Y F R., "The Growing Pains of Cloud Storage", IEEE Internet Computing, vol.19,no.1, pp. 4-7, 2015.
- [2] Zeng W, Zhao Y, Ou K, et al., "Research on cloud storage architecture and key technologies, International Conference on Interaction Sciences: Information Technology", Culture and Human 2009, Seoul, Korea, November 24-26, 2009, pp. 1044-1048.
- [3] Bocchi E, Drago I, Mellia M., "Personal cloud storage: Usage, performance and impact of terminals", IEEE, International Conference on Cloud NETWORKING, IEEE, 2015.
- [4] Wang C, Chow S S M, Wang Q, et al., "Privacy-Preserving Public Auditing for Secure Cloud Storage", IEEE Transactions on Computers, vol.62, no.2, pp.362-375, 2013.

- [5] Wu J, Ping L, Ge X, et al., "Cloud Storage as the Infrastructure of Cloud Computing", International Conference on Intelligent Computing and Cognitive Informatics, 2010, pp. 380-383.
- [6] Spillner J, Müller J, Schill A., "Creating optimal cloud storage systems", Future Generation Computer Systems, vol.29, no.4, pp.1062-1072, 2013.
- [7] Bonvin N, Papaioannou T G, Aberer K., "A self-organized, fault-tolerant and scalable replication scheme for cloud storage", Proc. of the ACM Symposium on Cloud Computing 2010 (SOCC2010), ACM, 2010, pp. 205-216.
- [8] Hu W, Yang T, Matthews J N., "The good, the bad and the ugly of consumer cloud storage", Acm Sigops Operating Systems Review, vol.44, no.3, pp. 110-115, 2010.
- [9] Corradi A, Fanelli M, Foschini L., "VM consolidation: A real case based on OpenStack Cloud", Future Generation Computer Systems, vol.32, no.1, pp. 118-127, 2014.
- [10] Palankar M R, Iamnitchi A, Ripeanu M, et al., "Amazon S3 for science grids: a viable solution?" Pp", Faculty Publications, pp. 55-64, 2015.
- [11] Factor M, Meth K, Naor D, et al., "Object storage: the future building block for storage systems", Local To Global Data Interoperability - Challenges and Technologies, 2005, pp. 119-123.
- [12] Zhang Y, Wang D., "Research on object-storage-based intrusion detection", vol.1, no.1, pp. 68-78, 2006.
- [13] Shapiro M, Mosseri L., "A Simple Object Storage System", Persistent Object Systems, Proceedings of the Third International Workshop, January 10-13 1989, Newcastle, New South Wales, 1989, pp. 272-276.
- [14] Kim J M, Jeong H Y, Cho I, et al., "A secure smart-work service model based OpenStack for Cloud computing", Cluster Computing, 2013, vol.17, no.3, pp.691-702.
- [15] Amazon B., "Amazon s3 service level agreement", 2012.